

Poster presentation

Finding new potential acetylcholine esterase Inhibitors in SDFiles using CWM Lead Finder and PASS (Prediction of Activity Spectra for Substances)

Hans-Jürgen Himmler* and Alexander Kos

Address: AKos Consulting & Solutions Deutschland GmbH (AKos GmbH), Austr. 26, D-79585 Steinen, Germany

* Corresponding author

from 3rd German Conference on Chemoinformatics
Goslar, Germany. 11-13 November 2007

Published: 26 March 2008

Chemistry Central Journal 2008, 2(Suppl 1):P42 doi:10.1186/1752-153X-2-S1-P42

This abstract is available from: <http://www.journal.chemistrycentral.com/content/2/S1/P42>

© 2008 Himmler and Kos

Chemical Workflow Manager (CWM) provides the non computer specialist a graphical user interface (GUI) to execute predefined chemistry related workflows. We provide several such predefined (and supported) workflows as products such as CWM Lead Finder and CWM Structure Comparer. CWM products are ready to use 'one-click' workflows. CWM is not a toolbox such as most of other workflow shells. Our vision is to incorporate free scientific software (such as CDK [1], Weka [2] and others) in ready to use and supported products usable by non computer specialists using a common user interface.

In the poster we present CWM Lead Finder. CWM Lead Finder reads a (in most cases rather small) SDFFile containing structures with known biological activity and reads a (normally rather large) SDFFile with structures of unknown activity. During execution of the workflow, the PASS (Prediction of Activity Spectra of Substances) program [3] is used to calculate prediction coefficients for around 3000 biological activities for both the structures with known and unknown' activity. These numbers are the difference between the percentage of prediction of being active (Pa) and being inactive (Pi). Using a cutoff factor the user can create a 'biological profile' (biological effects predicted with high probability for structures with known activity) that is further used in the workflow. Using this activity list the prediction coefficients are calculated for both the structures with known and unknown activity and afterwards are clustered. Structures from the SDFFile with unknown activity that end up in clusters that

contain structures with known activity have a high probability to show the same biological effects, thus are potential candidates for testing against this activity. This application answers the question: "Which compounds are active?"

The workflow goes one step further, and helps to present the most interesting compounds. The workflow selects compounds from clusters that have the same biological profile as the 'known' compounds, and calculates the chemical fingerprints. We use the CDK ² for calculating the chemical fingerprints. The fingerprints are again clustered: some clusters are a mixture of compounds of known and unknown activity, some contain only 'known' compounds, and other only 'unknowns'. The most interesting clusters to look at are the ones that contain exclusively structures with unknown activity, since those structures are divers to the structures of known activity but have a high probability to show the same activity.

The aim of CWM Lead Finder is to allow users that are not computer specialists or do not have in-depth know-how about prediction software to easily test their compounds for biological activity. Chemical Workflow Manager products are always predefined workflows that are (in the easiest case) 'one-click' workflows. For advanced users the CWM user interface also allows to define new workflows, respectively modifying existing workflows.

Chemical Workflow Manager is based on the new Microsoft Workflow Foundation (WF) [4]. WF is part of .Net and therefore available for free. It provides a framework for developing workflow centric solutions.

Workflows executed in Chemical Workflow Manager are tightly integrated in Microsoft SQL Server, thus allowing to run workflows from the database (i.e. eliminates the problem of local files). All relevant information about the executed workflow is stored in the SQL Server database allowing easy generation of reports about how the results got obtained.

References

1. . [<http://sourceforge.net/projects/cdk>]
2. Witten Ian H., Frank Eibe: **"Data Mining: Practical machine learning tools and techniques"**. 2nd Edition edition. Morgan Kaufmann, San Francisco; 2005. (2005) [<http://www.cs.waikato.ac.nz/ml/weka/>]
3. Poroikov VV, Filimonov DA, Borodina YV, Lagunin AA, Kos A: **Robustness of biological activity spectra predicting by computer program PASS for non-congeneric sets of chemical compounds.** *J. Chem Inform Comput Sci* 2000, **40(6)**:1349-1355. [<http://www.akosgmbh.eu/pass>]
4. . [<http://msdn2.microsoft.com/en-us/netframework/aa663328.aspx>]

Publish with **Chemistry**Central and every scientist can read your work free of charge

"Open access provides opportunities to our colleagues in other parts of the globe, by allowing anyone to view the content free of charge."

W. Jeffery Hurst, The Hershey Company.

- available free of charge to the entire scientific community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
<http://www.chemistrycentral.com/manuscript/>



ChemistryCentral